8.5 Appendix: Normal Score Plots in R

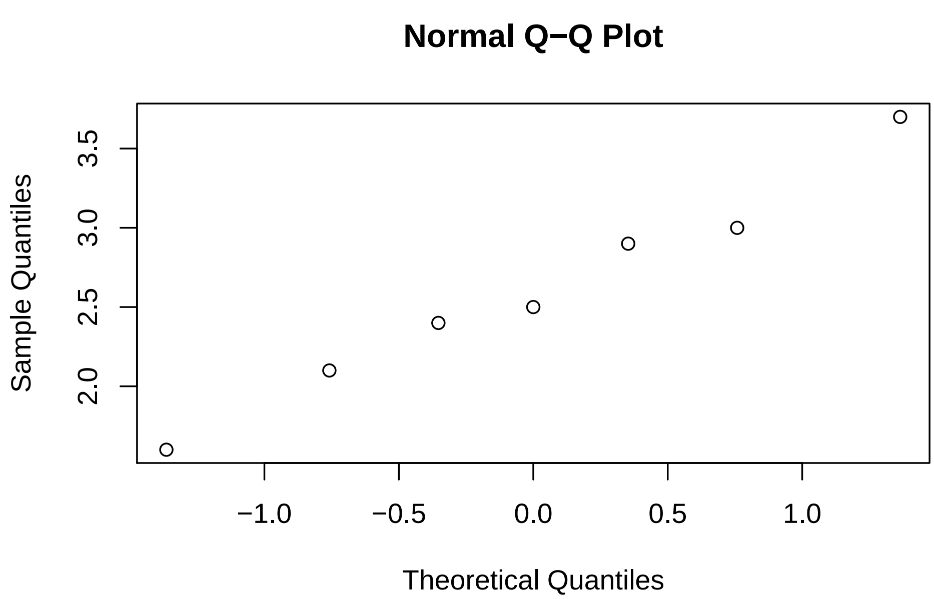
by EV Nordheim, MK Clayton & BS Yandell, September 9, 2003

In this chapter we have used both Minitab and R to construct normal scores plots. In R this is done using the function `qqnorm`, which automatically generates the plot. Before we show its use, we have three important points to note regarding normal scores plots made with computer packages. First, different packages handle ties in different ways. The different approaches result in similar-looking plots, however, so we will not dwell on this point. More noteworthy is the fact that some packages reverse the axes: they plot the population quantiles on the vertical axis and the sample quantiles on the horizontal axis. In these cases, the slope and intercept for such a plot have different meanings than for the plots we have described; however, the key point remains the same: data from a normal population should result in a plot where the points follow a straight line.

Our third point is that, instead of using the population quantiles $z_{[(i-0.5)/n]}$ we discussed earlier, some computer packages use slight variants on this. For example, R uses $z_{[(i-3/8)/(n+1/4)]}$ for small sample sizes. The reasons for these variants are somewhat technical and do not concern us here. The differences in the resulting plots are very small.

With these points in mind, we now describe the construction of normal scores plots in R. Consider some artificial data called `mydata`. To produce a normal scores plot, you type:

```
> mydata = c(2.4, 3.7, 2.1, 3, 1.6, 2.5, 2.9)
> myquant = qqnorm(mydata)
```



If the observations in `mydata` come from a normal distribution, then the above plot of `mydata` versus their population quantiles should give a straight line. It seems not unreasonable to conclude from this plot that the data come from a normal distribution.

The object `myquant` contains the quantiles (`myquant$x`) and the original data (`myquant$y`). Note that 1.6 is the smallest value in the data set. The corresponding population quantile $z_{[(i-3/8)/(n+1/4)]}$ from a standard normal distribution where $i = 1$ and $n = 7$. (You should check that $P(Z \leq -1.36) = 0.0869 \doteq 0.0862 = (1 - 3/8)/(7 + 1/4)$.) The quantiles can be viewed by printing the object `myquant` as a data frame:

```
> data.frame(myquant)

           x   y
1 -0.3529340 2.4
2  1.3644887 3.7
3 -0.7582926 2.1
4  0.7582926 3.0
5 -1.3644887 1.6
6  0.0000000 2.5
7  0.3529340 2.9
```